第4卷 第5期
1993年10月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 4, No. 5
Oct. 1993

解释学习进展

石纯一　　黄毅青

(清华大学计算机系,北京 100084)

EXPLANATION—BASED LEARNING AND ITS PROGRESS

Shi Chunyi and Huang Yiqing

(*Computer Science Department*, *Tsinghua University*, *Beijing* 100084)

Abstract This paper describes recent development in explanation-based learning (EBL), including its mechanism; several important up-dated research issues on EBL, such as formalization of operationality, typical EBL procedure explanation and genelization; straleziestowards domain theory problems, also the auther has proposed some ideas on characters. overall development and continuing research direction for EBL.

摘要 本文对解释学习近期的研究进行了论述,包括解释学习机制,当前解释学习研究的几个重要问题:如可操作性的形式化定义、典型解释学习过程的解释与概括、针对不完善领域理论的多种解决办法等,还就解释学习特点、总体发展状况及今后研究方向提出了看法.

§0. 引　言

机器学习是机器智能研究的重要方面,至今有不少的机器学习方法,包括从大量例子中获取知识,如归纳学习,类比学习;也有从单个例子中学习的方法,本文所讨论的解释学习(Explanation-Based Learning)方法属于后者. EBL 最早可追述到70年代初的机器人规划系统 STRIPS 中形成宏操作技术,80年代初 DeJong, Silver. Mitchell 等进一步提出解释学习的概念,这之后 AI 研究者进行了积极的研究,如今解释学习已成为机器学习方法的一个活跃的分支,并在实际应用中显示出它的威力.

EBL 以要学习的目标概念的某个具体例子为输入,运用领域知识解释例子是否为目标概念的真正实例,而后概括该解释来获取关于目标概念的一般性描述或解决问题的一般方法等知识,纵观 EBL 的发展过程,特别是近期的研究动态,可看出解释学习方法的几个特点:

(1)运用解释学习方法的学习器必须具有领域理论、目标概念、可操作性准则的起码知

本文1991年6月收到. 本文是国家自然科学基金资助项目. 作者**石纯一**,教授,主要研究领域为人工智能应用基础,知识工程. **黄毅青**,1991年毕业于清华大学,现在美国读研究生,主要研究领域为人工智能.

© 中国科学院软件研究所 http://www.jos.org.cn

识.

(2)从不同例子中学习的结果会使同一个目标概念的描述多种多样.

(3)解释学习并非真正的知识获取形学习方法.它是将目标概念的原有不可操作描述转化为可操作描述,使系统的运行效率更高.

(4)解释学习是从单个例子中学习,因此为了证明例子是目标概念的训练例必须借助领域知识.

(5)一般而言,解释学习可以学习任何它可解释的现象,既可以是正面例子,也可以是反面例子.大部分系统都从成功例中进行概括,也有系统从失败例中寻找失败的原因,从而同样达到学习的效果.

§1.解释学习机制

1.1 基于解释的概括学习机制(EBG)

1986年Rutgers大学的T.M.Mitchell等人给出了一种解释学习模型——基于解释的概括(EBG),它较成功地刻划了解释学习机制,并指出了解释学习进一步研究的问题.

为从单个训练例中形成可信赖的概括规则,必须赋予学习系统以解释能力,其中特别重要的是解释该训练例为什么是要学习的目标概念的正确实例,从解释中得到目标概念的概括性描述,概括的前提是运用领域知识对要学习的概念进行解释,这正是EBG的出发点.

为理解EBG方法,Mitchell先定义了概念、实例、概念描述、概括、解释等概念,然后给出EBG中要解决问题的描述,最后阐述了EBG方法及其具体实现技术.

EBG采用逻辑描述知识,并以推理观点看待求解,强调的是目标概念及其描述的转化.一阶谓词逻辑是其形式化基础,在这种描述方法下,对可操作性的判别成为可能,而且EBG也具有了比较明确的语义.EBG中之所以将目标视为概念,通过学习例子将概念描述转化为可操作的描述等都是由于EBG是一种逻辑理论构筑的框架.

1.2 别于EBG的学习机制

1986年R.Mooney等人给出了EGGS算法,它也是一种描述EBL的学习机制.EGGS即利用全局替换的解释概括,虽然它采用的具体实现技术与EBG不同,但其基本思想是一致的.

有意思的是1986年DeJong从与EBG不太相同的方向提出了更广泛的解释学习概念,DeJong和Mitchell不同想法著成的论文登于同一本《机器学习》杂志,DeJong从问题求解及其问题空间观点来看待EBL,强调的是问题状态和获取初始状态转化为目标状态的算子、规划等的方法,概括成了获取解决问题方法(原文Schemeta:框架之类的知识包)的过程.这里关键是通过EBL用区别于产生式规则的更一般的知识表达方式Schemeta来解决更复杂领域,如自然语言理解、规划方面的问题.

DeJong基于上述观点对EBL问题又给出了一套别于EBL的描述,而且在相应的工作原理中也体现了别于EBG的地方.解释过程既得到了解释,又得到了概括后的解释结构,虽然这时没对可操作性进行处理,而是在之后的概括过程中才真正得到依可操作性准则修剪的概括结构.另外,他给出的可操作性准则不是输入系统的静态定义,而是基于证明过程的动态定义.

© 中国科学院软件研究所 http://www.jos.org.cn

1987 年 Haym Hirsh 在逻辑程序环境下实现了 EBG 方法,但他更重要的工作在于将解释与概括合为一体,证明训练例为目标实例时就产生了概括结构.

Haym Hirsh 在上述思想的指导下,重新认识 Vllmen 的 MRS 系统. MRS 系统提供了许多推理形式,其基本表示方法就是逻辑的,用户给系统提供一系列的规则,选择证明事实的推理形式. 这样,逻辑规律给 EBG 以清晰的语义,并允许 EBG 与其它的逻辑方法结合起来. 另外,在逻辑意义下可动态推理可操作性.

§2. 解释学习研究的主要问题

2.1 可操作性的定义

解释学习是转化目标概念的初始描述为最终描述的知识转化的学习,其中区分最终学到的描述与系统给出的初始描述的关键在于得到的描述是否满足可操作性准则. 虽然从解释学习提出之日起就意识到可操作性及其判别准则的重要性,却一直没有给出可操作性的精确定义,只有含糊的阐述,因此不同系统实现中有各自不同的处理方法.

一般认定学习得到对系统来说好用的规则就是满足可操作性准则的规则,该定义在 DeJong,Mitchell,Mooney 等人的文章中都曾提到,根据各自不同的研究内容,对可操作性的具体描述又互不相同.

Mitchell 认为可操作性准则是一些描述例子的谓词,因为一般这些谓词当系统建立时就已存在,所以只是一种静态定义;DeJong 认为可操作的概念描述是指其真实性容易被证实;可操作性判断在他看来是上下文有关的,在一种情况下可操作的规则在另一种情况下不一定可操作,只有在给定的环境状态下可操作性判断才成为可能;DeJong 特别强调可操作性动态定义,认为可操作性判断可看作是一个证明结构的函数;另一种动态定义的方法是应用定理证明器来动态决定可操作性. 所列举的这些关于可操作性的定义方法虽不完善,尚未形式化,但都有可取之处.

基于前面的工作,Stanford 大学的 R. M. Keller 作了较深入的探索. 虽然他的可操作性描述仍然只是文字的阐述,但他的工作为统一可操作性的一般定义、明确可操作性基本内含,估价测试可操作性等一贯含糊的问题的解决奠定了良好的基础.

Keller 首先引入三个概念:"概念"、"概念描述"、"例子"并相应给出概念空间、概念描述空间、例子空间. 可操作性可看作是概念描述空间的搜索,搜索结果是将不可操作的描述转为可操作的描述. Keller 特别指出概念与概念描述的区别,这是他与将可操作性看作是搜索过程观点的 Mitchell 的主要不同之处.

Keller 的定义清晰地指出了可操作性是什么,而且作者进一步认为在不同的系统中可操作性的估价标准是不同的,同一系统中可操作性随着时间的推移也是变化的,所以有静态动态之分,这基本解决了原来不同系统关于可操作性的差异,并可给出估价标准.

2.2 解释过程

解释就是依照领域理论给出训练例符合目标概念要求的证明. 解释结构建立后每个分支上的叶节点都必须满足可操作性准则. 解释是解释学习的第一步,解释后才可能概括. 建立解释涉及如下几个重要问题.

(1)必须有充足的领域知识来保证,解释结点的扩展以知识库规则为前提,要求最后到

© 中国科学院软件研究所　http://www.jos.org.cn

达的叶节点可操作.解释的方法在很大程度上受领域知识形式和内容的影响,给建立解释造成许多困难.

(2)因为满足每个解释结点的规则可能有多条,对一个例子的解释往往有多个,这样会导致学到不同的规则,问题在于多条路径证明中并非每条都能满足可操作性要求,因此如何选择解释结点处的规则是非常重要的.

(3)解释过程在依据充足领域知识前提下,还有两种可能,一种是直接运用知识,另一种是由系统内部借助观察或分析别的问题解决方法来达到目的,两种途径需要不同层次的问题解决能力.

2.3 概括过程

概括是解释学习的核心,Mitchell 提出的 EBG 学习机制的关键部分就是概括.1987 年 DeJong 将解释学习中的概括分为以下四类:

(1)无关特性删除:删除解释过程中没有用到的训练例的特性.

(2)个性删除:去掉对特定物体不必要的依赖性.

(3)可操作性修剪:去掉解释中容易重建的子解释结构.

(4)结构化概括:改变解释的内部结构,它包括重组,转换和向解释中加入新成分,如选言扩充、时态概括和数字概括.

EBG 只能做前两种概括,局限性很大,更重要也更难实现的是(3)、(4)两种概括.

总的来说,概括涉及如下问题:

(1)何时进行概括?

DeJong 研究自然语言理解的系统时,对每个学习的目标从目标是否一般等五个方面考察后,决定是否应该概括;P. M. Andress 提出必须对概括加以限制以及约束概括的三条原则;Minton 指出盲目学习会使知识无限制地增加,系统必须具有进行选择性学习的机制.

(2)概括与解释的结合

前述介绍不同的解释学习机制时已涉及三种处理解释与概括的结合过程,其中较好的当推将解释与概括合二为一的过程.

(3)解释学习的概括得到的目标概念描述往往是目标概念所代表的外延区域的子集,即过于细化,也有少数情况下概括过头,该问题的判定是不可解的.

(4)概括的具体技术

(5)除了关于概括的一般框架研究,另外很重要的研究方向是将概括机制揉进执行系统作为系统的组成部分来提高系统的执行效率.

2.4 解释学习后结果的整理

2.4.1 效用性问题

解释学习的效用性问题是指系统通过学习提高系统执行效率的问题,学习的最终目的是为了提高系统的运行效率,解释学习也不例外.1988 年 Minton 研究指出解释学习不一定能达到效用性要求,他是用 PRODIGY 系统来说明所学到的控制知识并不都有用,只有当运用所学知识给系统带来的效益超过系统测试所学知识是否可用的耗费量时,所学到的知识才算达到了效用性目标.Mooney 从学到的信息的使用效果发现实验及非形式化的分析都表明无限制地使用学到的规则最后可能导致执行系统效率降低;相反,限制使用能去除学

© 中国科学院软件研究所 http://www.jos.org.cn

到的规则引起的反作用提高系统性能. 这说明如何使用学到的规则会极大地影响这些规则最终是提高还是降低整体执行效果.

2.4.2 概念求精

学习往往不是一次就能学到新知识或有用的知识，EBL 是从单个例子中学习就更加存在多次学习的问题. 不断改进已学到知识的过程就是概念求精. 很多研究者包括 DeJong 都非常重视概念求精的问题，但由于求精过程涉及面较广，研究起来并非易事，这方面的工作还不够深入.

在求精研究中值得一提的是增量逼近的方法. 由于初始学到的知识往往不准确或不完善有必要进行修正；通过观察系统对该知识或类似知识学习的结果可修正原有的初始知识，这种观察往往是基于初始知识运用时发生了错误而激发的. S. A. Chien 正是从这方面着手，遁着规划时的事件链逐步找到初始知识并对其求精，该过程是增量逼近，所以称之为增量逼近法.

2.4.3 粒度问题

粒度是 Rutgers 大学的 S. Roy1988 年在研究 EBL 时，学习后的规则给知识库造成冗余的问题而提出来的. Roy 认为由于训练例规模的不同，EBL 会获得大小各异的规则，该互异性被称之为粒度. 粒度造成了学到规则的冗余，缩小了规则的应用范围. 针对粒度，他提出"综合分析"法来减小学到规则的粒度，提高学到规则的适用面. 电路设计系统中粒度问题特别突出，Roy 选择 VEMED 系统的学习器 LEAP 作为综合分析法的实现对象. 该法是针对 LEAP 学到的大粒度规则，从手工求精过程中得到尽可能细的规则，具体讲就是考察手工的求精步骤，从中抽取可用现存规则来解释的部分概括为一条新规则.

对于粒度，Roy 并没有给出明确的定义，只是对粒度粗细进行了相对描述，规则 r 的粒度大于规则 r′的粒度是指可构造规则树 t 含有 r′，t 应用可产生出与 r 规则产生的一样的结果. 比如 rl 中包含了 r2、r3，r2、r3 可分别与不同的规则组合得到粒度更大的规则，因此 r2、r3 比 r1 适用面更广，但并非学到的规则粒度越小越好，理想的粒度是粗到可有效推理，细到可明显区分各条规则.

2.5 不完善领域理论

解释学习虽然得到广泛应用，但由于系统执行性能受到最初放入领域理论知识的极大限制，在知识库中汇集全部适当的知识又难于实现，势必造成不完善的领域理论，给解释学习的进行带来困难. DeJong 及 Mitchell 等人将不完善领域理论分为三大类：

①不完备的理论. 这是由于缺少一些公理或规则，使某些演绎不能完成.

②不一致的理论. 它可导出不一致的结论.

③难处理的理论. 演绎由于受到计算复杂性的限制而不能完成.

针对这三类问题许多研究者进行了大量工作，很多文献探讨了这方面的问题.

§3. 解释学习系统

DeJong、Mitchell、Mooney Shavlik 等均建立过很出色的基于解释的学习系统. 一般从学习系统是否需要外界知识的角度可把解释学习系统分为三种类型：

(1)从师型学习系统. 之所以称之为从师是指当遇到疑难及系统无法解决的问题时有教

© 中国科学院软件研究所　http://www.jos.org.cn

师或专家指点，从外部输入专家的解决方法，解释该方法从中学习. 这类系统有 Minton 等人设计的 PRODIGY 学习系统，Mitchell 等构筑的学习器 LEAP，P. O. Rocke 的 MA 系统.

(2)观察自身以外行为来学习. 区别于从师型系统的关键是它能在无教师指导的情况下观察外界解决问题的方法，进行解释后学习，这类系统的任务之一是决定哪些例子及其性质是重要且值得学习的，虽然同样可以是学习专家的知识，但系统不会打扰专家的工作，只是观察，从中选择性地学习. 这类系统有 Ellman 设计系统的学习器，Minton 的下棋规划系统，Mooney、DeJong 的自然语言理解系统.

(3)无需外部指导者. 其基础是完全依赖知识库，DeJong 在这方面作了很有意思的尝试.

为清楚地了解解释学习方法的实际系统，下面简单说说几个比较典型的系统.

(1)PRODIGY

PRODIGY 是从机器人规划系统 STRIPS 扩展而来的，主要用于学习控制知识，它的三个组成元素是细化、压缩和效用评估. 值得一提的是 PRODIGY 用的并非标准的 EBG 算法，而是用 EBS(Explanation-Based Specification)算法来解释为什么例子满足目标概念描述，其中并没有建立明确的解释结构而是用 EBS 算法递归地扩展一个非明确的解释的最弱前提条件来达到解释的目的.

(2)LEAP

Mitchell 1985 年为 VLSI 电路辅助设计系统 VEXED 建立了一个学习器 LEAP，LEAP 是一种从师型学习系统，当遇到系统无法设计的电路时，专家给出解法，LEAP 从中学习，得到新的设计方法.

(3)PHYSICS 101 系统

该系统是 Shavlik 和 DeJong1990 年在数学方法领域，针对大学程度的物理问题，运用解释学习方法设计的学习系统. 系统包括学习理论及要解决的问题，它通过分析老师解决问题的特殊方法，概括这些方法来提高自己解决问题的能力. 该系统的概括是针对我们前述提到的结构化概括中的数字概括，从待定数目的实例概括出一般数字时的规律，Shavlik 的工作在数字概括方面是相当突出的. 另外 PHYSICS 101 比较好地解决了可操作性与概括性的相互关系.

(4)SOAR 系统

SOAR 虽然不是典型的解释学习系统，但从认知观点看是个很典型的系统. 它对一般概括过程的形式化进行了很好的尝试. SOAR 是问题空间的启发式搜索来形式化所有行为(包括问题和任务)的问题求解系统，另一显著特点是采用知识块(CHUNK)来表达知识.

EBL 学习系统还有一些，在此不一一赘述.

§4. 近期的解释学习研究方向

(1)对于可操作性的全面认识仍然很模糊，它是 EBL 的中心问题，很重要的一点是可操作性随着周围世界而变化，因此其定义也随之而变化，是动态的.

(2)形式化工作非常重要. 比如，概括研究中的结构化概括是有趣而困难的，我们希望给出一个适于任何领域的一般结构化概括方法，但遗憾的是这办不到，结构化概括必须依靠领

© 中国科学院软件研究所 http://www.jos.org.cn

域知识的特殊性.

(3)EBL 不是机器学习的完美答案,需要做大量的工作将解释学习与其它学习方法结合起来,比如,与类比学习的结合.

(4)正如人的学习过程需要反复,几乎所有领域的学习都需要求精的过程,解释学习的求精研究也需要花大力气.

(5)不完善领域问题一直是解释学习的主要难点,许多领域相对机器描述太复杂,领域中无法完全包括所有需要的知识.

(6)EBL 概括何时该进行值得很好研究,现在对概括本身已作了大量研究,但概括与系统本身运行的结合没有很彻底的研究.

(7)解释学习的心理学基础应该认真探讨.一般的心理学方面的知识获取仅仅包括智能的概念,但近期,许多心理学家开始怀疑上下文无关的关于学习的理解.另外,许多事实证明 EBL 从心理学角度分析很有效,正如在 SOAR 系统中所体现出来的那样.

参考文献

1 G. DeJong, An Introduction to Explanation-Based Learning, AAAI-88 Tutorial.

2 S. Minton, Learning Search Contral Knowledge, Chap4-6.

3 R. Barletta, Explanation-Based Lndering of Cases, AAAI-89.

4 G. DeJong, Explanation-Based Learning: an Alternative View, Machine Learning ,1. 1986.

5 T. M. Mitchell, Explanation-Based Learning: an Onifying View, Machine Learning, 1. 1986.

6 J. W. Shavlik, Acquiring Recursive and Iterative Conceptswith EBL, Machine Learning, 5. 1990.

7 R. Mooney, The Effect of Rule Use on the Utility of Explanation-Based Learning, IJCAI-89.

8 S. J. Russell, IMEX: Overcoming Intraclality in Explanation-Based Learning, AAAI-88.

9 S. W. Bennat, Approximation in Mathematical Domains, IJCAI-87.

10 S. Minton, Selectivety Generalizing Plans for Problem-Solving, IJCAI-87.

11 K. Mikeller, Defining Operationality for Explanation-Based Learning, Artificial Intelligence 34.

12 A. M. Seyre, On the Operationality/Generality Trade-off in Explanation-Based Learning, IJCAI-87.

13 D. Kibter, Perturbation: a Means for Guiding Generalization, IJCAI-87.

14 T. E. Laird, Towards Chuaking as a General Learning Mechanism, AAAI-84.

15 S. Roy, Parsing to Learn Fine Grained Rules, AAAI-88.

16 S. Razamony, The Ctassification, Detection and Handling of Imperfect Theory Problems, IJCAI-87.

17 T. Euman, Approximate Theory Formation: an Explanation-Based Approach, AAAI-88.

18 N. S. Flann, A Study of Explanation-Based Methods for Inductive Learning, Machine Learning, 1990.

19 石纯一、龚义涛,"解释学习的可操作性",计算机学报,1992. 2.

20 石纯一、黄毅青,"解释学习中的求精算法及其实现",知识工程进展,1991.

© 中国科学院软件研究所　http://www.jos.org.cn